JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

ELSEVIER Journal of Computational and Applied Mathematics 124 (2000) 25–44

The theory of Newton's method

A. Galántai

Institute of Mathematics, The University of Miskolc, 3515 Miskolc-Egyetemvaros, Hungary

Received 18 June 1999; received in revised form 31 January 2000

Abstract

We review the most important theoretical results on Newton's method concerning the convergence properties, the error estimates, the numerical stability and the computational complexity of the algorithm. We deal with the convergence for smooth and nonsmooth equations, underdetermined equations, and equations with singular Jacobians. Only those extensions of the Newton method are investigated, where a generalized derivative and or a generalized inverse is used. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

The Newton or Newton–Raphson method has the form

$$x_{k+1} = x_k - [F'(x_k)]^{-1}F(x_k), \quad k = 0, 1, \ldots \tag{1}$$

for the solution of the nonlinear equation

$$F(x) = 0 \quad (F : X \to Y), \tag{2}$$

where X and Y are Banach spaces and F' is the Fréchet derivative of F. The geometric interpretation of the Newton method is well known, if F is a real function. In such a case x_{k+1} is the point where the tangential line $y - F(x_k) = F'(x_k)(x - x_k)$ of function $F(x)$ at point $(x_k, F(x_k))$ intersects the x-axis. The geometric interpretation of the complex Newton method ($F : C \to C$) is given by Yau and Ben-Israel [70]. In the general case $F(x)$ is approximated at point x_k as

$$F(x) \approx L_k(x) = F(x_k) + F'(x_k)(x - x_k). \tag{3}$$

The zero of $L_k(x) = 0$ defines the new approximation x_{k+1}.

Variants of the Newton method are the damped Newton method

$$x_{k+1} = x_k - t_k[F'(x_k)]^{-1}F(x_k) \quad (t_k > 0,\ k = 0, 1, 2, \ldots) \tag{4}$$

E-mail address: matgal@gold.uni-miskolc.hu (A. Galántai).

0377-0427/00/$ - see front matter
PII: S0377-0427(00)00435-0

and the modified Newton method

$$x_{k+1} = x_k - [F'(x_{\alpha_k})]^{-1}F(x_k) \quad (k = 0, 1, 2, \ldots), \tag{5}$$

where $\alpha_0 = 0$, $\alpha_{k-1} \leqslant \alpha_k \leqslant k$ $(k \geqslant 1)$. The latter formulation includes the Shamanskii–Newton method [42], where the Jacobian is always reevaluated after m consecutive iterations. The Newton-like methods are generally defined by the recursion

$$x_{k+1} = x_k - [M(x_k)]^{-1}F(x_k) \quad (k = 0, 1, 2, \ldots),$$

where $M(x)$ is usually an approximation to $F'(x^*)$, where x^* is a solution of Eq. (2). These methods formally include the quasi-Newton and inexact Newton methods, as well. In this paper we deal only with the theory of Newton's method. We concentrate on the convergence properties, error estimates, complexity and related issues. It is remarkable that much of these results were obtained in the last 30 years. Yet, the theory of Newton method is far from being complete. For the implementation of Newton's method we refer to Ortega–Rheinboldt [42], Dennis and Schnabel [13], Brown and Saad [8], and Kelley [29]. Kearfott [1, pp. 337–357] discusses the implementation of Newton's method in interval arithmetic. For other important results not quoted here we refer to the reference list of the paper.

2. Convergence results for smooth equations

Let X and Y be two Banach spaces. Let $S(x, r) = \{z \in X \mid \|z - x\| < r\}$ denote the open ball in X with center x and radius r and let $\overline{S(x, r)}$ be its closure. Let $F : S(x_0, R) \subset X \to Y$ be a given (nonlinear) mapping. Assume that a sphere $S(x_0, r)$ exists such that $\overline{S(x_0, r)} \subset S(x_0, R)$. Denote by $F'(x)$ and $F''(x)$ the first and second derivatives of F in the sense of Fréchet. Kantorovich proved the following classical result.

Theorem 1 (Kantorovich [27]). *Let $F : S(x_0, R) \subset X \to Y$ have a continuous second Fréchet derivative in $\overline{S(x_0, r)}$. Moreover, let* (i) *the linear operation $\Gamma_0 = [F'(x_0)]^{-1}$ exist*; (ii) $\|\Gamma_0 F(x_0)\| \leqslant \eta$; (iii) $\|\Gamma_0 F''(x)\| \leqslant K$ $(x \in \overline{S(x_0, r)})$. *Now, if*

$$h = K\eta \leqslant \tfrac{1}{2}$$

and

$$r \geqslant r_0 = \frac{1 - \sqrt{1 - 2h}}{h}\eta,$$

Eq. (2) *will have a solution x^* to which the Newton method is convergent. Here,*

$$\|x^* - x_0\| \leqslant r_0.$$

Furthermore, if for $h < \frac{1}{2}$

$$r < r_1 = \frac{1 + \sqrt{1 - 2h}}{h}\eta,$$

or for $h = \frac{1}{2}$

$$r \leqslant r_1,$$

the solution x^* will be unique in the sphere $\overline{S(x_0,r)}$. The speed of convergence is characterized by the inequality

$$\|x^* - x_k\| \leqslant \frac{1}{2^k}(2h)^{2^k}\frac{\eta}{h} \quad (k = 0, 1, 2, \ldots).$$

Remark 2. The conditions (ii) and (iii) of the theorem can be replaced by (i′) $\|\Gamma_0\| \leqslant B'$; (ii′) $\|F(x_0)\| \leqslant \eta'$; (iii′) $\|F''(x)\| \leqslant K'$ $(x \in \overline{S(x_0,r)})$. In this case h, r_0 and r_1 are, respectively,

$$h = K'(B')^2\eta', \quad r_0 = \frac{1-\sqrt{1-2h}}{h}B'\eta', \quad r_1 = \frac{1+\sqrt{1-2h}}{h}B'\eta'.$$

Remark 3. The conditions $h \leqslant \frac{1}{2}$ and $r_0 \leqslant r$ are necessary for the existence of solution. The solution is not unique in the absence of condition $r < r_1$ or $r \leqslant r_1$.

Notice that $\overline{S(x_0,r)}$ gives an inclusion region for a zero of Eq. (2). If a bound is known for $\|[F'(x)]^{-1}\|$ in $\overline{S(x_0,r)}$, the condition imposed on h can be weakened by requiring $h < 2$ instead of $h \leqslant 1/2$.

Theorem 4 (Mysovskikh [38]). *Let the following conditions be satisfied*: (i) $\|F(x_0)\| \leqslant \eta$; (ii) *the linear operation* $\Gamma(x) = [F'(x)]^{-1}$ *exist for* $x \in \overline{S(x_0,r)}$, *where* $\|\Gamma(x)\| \leqslant B$ $(x \in \overline{S(x_0,r)})$; (iii) $\|F''(x)\| \leqslant K$ $(x \in \overline{S(x_0,r)})$. *Then, if*

$$h = B^2K\eta < 2$$

and

$$r > r' = B\eta\sum_{j=0}^{\infty}\left(\frac{h}{2}\right)^{2^j-1},$$

Eq. (2) *has a solution* $x^* \in \overline{S(x_0,r)}$ *to which the Newton method with initial point* x_0 *is convergent. The speed of convergence is given by*

$$\|x^* - x_k\| \leqslant B\eta\frac{(h/2)^{2^k-1}}{1-(h/2)^{2^k}} \quad (k = 0, 1, 2, \ldots).$$

The Newton iterates x_k are invariant under any affine transformation $F \to G = AF$, where A denotes any bounded and bijective linear mapping from Y to any Banach space Z. This property is easily verified, since

$$[G'(x)]^{-1}G(x) = [F'(x)]^{-1}A^{-1}AF(x) = [F'(x)]^{-1}F(x).$$

The affine invariance property is clearly reflected in the Kantorovich theorem. For other affine invariant theorems we refer to Deuflhard and Heindl [14].

The Kantorovich theorem is a masterpiece not only by its sheer importance but by the original and powerful proof technique. The results of Kantorovich and his school initiated some very intensive research on the Newton and related methods. A great number of variants and extensions of his results emerged in the literature. Ortega–Rheinboldt [42] is a good survey for such developments until 1970 (see also [24,56] or [13]).

In the Newton–Kantorovich theorem the continuity conditions on $F''(x)$ can be easily replaced either by

$$\|F'(x)-F'(y)\|\leqslant K\|x-y\| \quad (x,y\in\overline{S(x_0,r)}) \tag{6}$$

or

$$\|[F'(x_0)]^{-1}(F'(x)-F'(y))\|\leqslant K\|x-y\| \quad (x,y\in\overline{S(x_0,r)}). \tag{7}$$

This was done by several authors, the first of which was, perhaps, Fenyő [16]. A typical result of this kind is the following.

Theorem 5 (Tapia [64]). *Let X and Y be Banach spaces and $F: D\subset X\to Y$. Suppose that on an open convex set $D_0\subset D$, F is Fréchet differentiable with*

$$\|F'(x)-F'(y)\|\leqslant K\|x-y\| \quad (x,y\in D_0).$$

Assume that $x_0\in D_0$ is such that (i) $[F'(x_0)]^{-1}$ *exists,*

$$\|[F'(x_0)^{-1}]\|\leqslant B,\quad \|[F'(x_0)]^{-1}F(x_0)\|\leqslant\eta,\quad h=BK\eta\leqslant\tfrac{1}{2},$$

(ii) $\overline{S(x_0,t^*)}\subset D_0$ $(t^*=((1-\sqrt{1-2h})/h)\eta)$. *Then the Newton iterates x_k are well defined, remain in $\overline{S(x_0,t^*)}$, and converge to $x^*\in\overline{S(x_0,t^*)}$ such that $F(x^*)=0$. In addition*

$$\|x^*-x_k\|\leqslant\frac{\eta}{h}\left(\frac{(1-\sqrt{1-2h})^{2^k}}{2^k}\right),\quad k=0,1,2,\ldots .$$

This result, which is often referred to as the Kantorovich theorem, is an improvement of Ortega [41] and gives an optimum estimate for the rate of convergence.

If the hypotheses (i) and (ii) of the above Kantorovich theorem are satisfied, then not only the Newton sequence $\{x_k\}$ exists and converges to a solution x^* but $[F'(x^*)]^{-1}$ exists in this case. The following result of Rall [50] shows that the existence of $[F'(x^*)]^{-1}$ conversely guarantees that the hypotheses of the Kantorovich theorem with $h<\frac{1}{2}$ are satisfied at each point of an open ball S^* with center x^*. In such a case x^* is called a simple zero of F.

Theorem 6 (Rall [50]). *If x^* is a simple zero of F, $\|[F'(x^*)]^{-1}\|\leqslant B^*$, and*

$$S_*=\left\{x\;\middle|\;\|x-x^*\|<\frac{1}{B^*K}\right\}\subset D_0,$$

then the hypotheses (i) *with $h<\frac{1}{2}$ and* (ii) *of the Kantorovich Theorem* 5 *are satisfied at each $x_0\in S^*$, where*

$$S^*=\left\{x\;\middle|\;\|x-x^*\|<\frac{2-\sqrt{2}}{2B^*K}\right\}.$$

The value given for the radius of S^* is the best possible.

Vertgeim [71] was the first to weaken the C^2 condition of $F(x)$ to the Hölder condition

$$\|F'(x)-F'(y)\|\leqslant K\|x-y\|^{\alpha} \quad (x,y\in D_0), \tag{8}$$

where $0 < \alpha \leqslant 1$ is a constant. His result was improved or rediscovered by several authors ([42,24,28, 55,3,4,33] and others). A typical result of this kind is the following

Theorem 7 (Jankó-Coroian [24]). *Let $F : S(x_0, R) \subset X \to Y$ be a given nonlinear mapping and assume that* (i) $\Gamma_0 = [F'(x_0)]^{-1}$ *exists and* $\|\Gamma_0\| \leqslant B$; (ii) $\|\Gamma_0 F(x_0)\| \leqslant \eta$; (iii)

$$\|F'(x) - F'(y)\| \leqslant K\|x - y\|^\alpha \quad (x, y \in \overline{S(x_0, r)},\ 0 < \alpha \leqslant 1),$$

where

$$r = \frac{(1+\alpha)^{1/\alpha}[(1+\alpha)^{1/\alpha}h]^{1-(1/\alpha)}}{(1+\alpha)^{(1/\alpha)} - 1}\eta,$$

(iv)

$$h = BK\eta^\alpha \leqslant \frac{\alpha}{1+\alpha}.$$

Then there is at least one zero x^* in $\overline{S(x_0, r)}$ and $\{x_k\}$ converges to x^* with the speed

$$\|x^* - x_k\| \leqslant \frac{[(1+\alpha)^{1/\alpha}h]^{(1+\alpha)^k - (1/\alpha)}}{[(1+\alpha)^{(1/\alpha)} - 1](1+\alpha)^{(k-1)/\alpha}}\eta.$$

Similar results hold for the Newton–Mysovskikh theorem (see [24]), one of which is the following.

Theorem 8 (Jankó [24]). *Let $F : X \to Y$ be a given nonlinear mapping and let the following conditions be satisfied*: (i) *the linear operator $\Gamma(x) = [F'(x)]^{-1}$ exists for all $x \in \overline{S(x_0, T\eta)}$, where*

$$T = \sum_{j=0}^{\infty}\left(\frac{h}{1+\alpha}\right)^{((1+\alpha)^j - 1)/\alpha}, \quad 0 < \alpha \leqslant 1,$$

(ii) $\|\Gamma(x_0)F(x_0)\| \leqslant \eta$;
(iii) $\|\Gamma(x)[F'(y) - F'(z)]\| \leqslant K\|x - y\|^\alpha \quad (x, y, z \in \overline{S(x_0, T\eta)})$;
(iv) $h = K\eta^\alpha < 1 + \alpha$.
Then Eq. (2) *has a solution $x^* \in \overline{S(x_0, T\eta)}$ to which the Newton method with initial point x_0 is convergent. The speed of convergence is given by*

$$\|x^* - x_k\| \leqslant T\eta\left(\frac{h}{1+\alpha}\right)^{((1+\alpha)^k - 1)/\alpha} \qquad (k = 0, 1, 2, \ldots).$$

Assuming that operator F is analytic, Smale [59], Rheinboldt [53], and Wang and Han [73] gave convergence results which utilize only information at the starting point. Denote by $F^{(j)}(x)$ the jth Fréchet derivative of F at point x. A typical result of this type is given in

Theorem 9 (Rheinboldt [53]). *Let $F : X \to Y$ be analytic on some open set S of X and let*

$$\beta(x) = \|[F'(x)]^{-1}F(x)\|, \quad \gamma(x) = \sup_{j>1}\left\|\frac{1}{j!}[F'(x)]^{-1}F^{(j)}(x)\right\|^{1/(j-1)}.$$

Consider a point x_0 of S where $F'(x_0)$ is invertible. Let ρ ($\approx$ 0.16842669) be the positive root of the cubic $(\sqrt{2}-1)(1-\rho)^3-\sqrt{2}\rho=0$. If $\alpha(x_0)=\beta(x_0)\gamma(x_0)\leqslant\rho/\sqrt{2}$ ($\approx$ 0.11909565) and the ball $S(x_0,r_0)$ with radius $r_0=\rho/\gamma(x_0)$ is contained in S, then the Newton iterates x_k converge to a solution x^* of Eq. (2)*. Moreover, the convergence is at least R-quadratic with R_2-factor $\frac{1}{2}$.*

Using the results of [73], Wang and Zhao [74] improved Smale's result by not assuming that $\gamma(x)$ is bounded.

3. Error estimates for the Newton method

There are several error estimates for the Newton method. The Kantorovich theorem is a basis for many of these, but there are others as well.

In the next four theorems we assume the conditions of Tapia's Theorem 5. Gragg and Tapia [20] added the following optimal error bounds to this theorem.

Theorem 10 (Gragg–Tapia [20]).

$$\|x^*-x_k\|\leqslant\begin{cases}\frac{2}{h}\sqrt{1-2h}\frac{\theta^{2^k}}{1-\theta^{2^k}}\|x_1-x_0\|, & \text{if } 2h<1,\\ 2^{1-k}\|x_1-x_0\|, & \text{if } 2h=1\end{cases}$$

and

$$\frac{2\|x_{k+1}-x_k\|}{1+\sqrt{1+4\theta^{2^k}/(1+\theta^{2^k})^2}}\leqslant\|x^*-x_k\|\leqslant\theta^{2^{k-1}}\|x_k-x_{k-1}\|,\quad k\geqslant 1,$$

where $\theta=(1-\sqrt{1-2h})/(1+\sqrt{1+2h})$.

Miel [34] gave a new proof for this theorem using the technique of Ortega [41]. An affine invariant version of the theorem was given by Deuflhard and Heindl [14]. Miel also constructed several error bounds for the Newton method [36,37].

Theorem 11 (Miel [36]). *Let us define the constants A_k, B_k and C_k recursively by*

$$A_1=\frac{1}{\eta}B_1,\quad A_{k+1}=A_k(2-\Delta A_k),\quad \Delta=\frac{2\eta\sqrt{1-2h}}{h},$$

$$B_1=\theta,\quad B_{k+1}=B_k^2,\quad \theta=(1-\sqrt{1-2h})/(1+\sqrt{1+2h}),$$

$$C_1=B_1,\quad C_{k+1}=\frac{C_k^2}{2C_k+\Delta/\eta}.$$

Then the error bounds

$$\|x^*-x_k\|\leqslant A_k\|x_k-x_{k-1}\|^2\leqslant B_k\|x_k-x_{k-1}\|\leqslant C_k\|x_1-x_0\|,$$

are valid and the best possible.

Theorem 12 (Miel [37]). *Let* $\Delta = (2\eta\sqrt{1-2h})/h$ *and*

$$\theta = (1-\sqrt{1-2h})/(1+\sqrt{1+2h}).$$

Then

$$\frac{2\|x_{k+1}-x_k\|}{1+\sqrt{1+(4/\Delta)(1-\theta^{2^k})/(1+\theta^{2^k})\|x_{k+1}-x_k\|}} \leqslant \|x^*-x_k\| \leqslant \frac{1-\theta^{2^k}}{\Delta}\|x_k-x_{k-1}\|^2$$

if $2h<1$, *and*

$$\frac{2\|x_{k+1}-x_k\|}{1+\sqrt{1+(2^k/\eta)\|x_{k+1}-x_k\|}} \leqslant \|x^*-x_k\| \leqslant \frac{2^{k-1}}{\eta}\|x_k-x_{k-1}\|^2$$

if $2h=1$.

Other proofs of this result can be found in [68,69]. Using nondiscrete mathematical induction, Pták [46], Potra and Pták [45], and Lai and Wu [31] gave convergence results and error estimates for the Newton and Newton-like methods. Here we recall the following result.

Theorem 13 (Potra–Pták [45]). *Let* $a = (\eta\sqrt{1-2h})/h$ *and*

$$\gamma(r) = (a^2+4r^2+4r(a^2+r^2)^{1/2})^{1/2} - (r+(a^2+r^2)^{1/2}).$$

Then

$$\gamma(\|x_{k+1}-x_k\|) \leqslant \|x_k-x^*\| \leqslant (a^2+\|x_k-x_{k-1}\|^2)^{1/2} - a.$$

Yamamoto [68] pointed out that the Gragg–Tapia estimates are derivations of the Kantorovich recurrence relations and the Potra–Pták and Miel estimates are improvements of the Gragg–Tapia theorem. He also showed that the latter two results also follow from the original Kantorovich theorem, and Miel's result (Theorem 12) improves on that of Potra and Pták. Yamamoto [69] gives a method for finding sharp posterior error bounds for Newton's method under the assumptions of Kantorovich's theorem. Yamamoto's paper [69], where a comparison of the best known bounds can also be found, is the best source for error estimates in the theory of the Newton method.

Other type of error estimate is given by Neumaier [40]. For any two vectors $x, y \in \mathbb{R}^n$ let $x \leqslant y$, if and only if $x_i \leqslant y_i$ for all i. Furthermore, let $|A| = [|a_{ij}|]_{i,j=1}^{m,n}$ for any $A \in \mathbb{R}^{m\times n}$.

Theorem 14 (Neumaier [40]). *Let* $F: D \subset \mathbb{R}^n \to \mathbb{R}^n$ *be continuous*, $x_0 \in D$, $A \in \mathbb{R}^{n\times n}$ *be nonsingular and* $\delta_0 = A^{-1}F(x_0)$. *Suppose further that there is a constant* $\kappa > 1$ *such that*

$$\overline{S(x_0, \kappa\|\delta_0\|)} \subset D$$

and a nonnegative vector $c \in \mathbb{R}^n$ *such that, for some monotone norm* $\|.\|$

$$|F(x) - F(x_0) - A(x-x_0)| \leqslant \|\delta_0\| c \quad (x \in \overline{S(x_0, \kappa\|\delta_0\|)}).$$

If the vector $b = |A^{-1}|c$ *satisfies the condition* $\|b\| \leqslant \kappa - 1$, *then* $F(x)$ *has at least one zero in* $\overline{S(x_0, \kappa\|\delta_0\|)}$, *and any such zero* $\hat{x}$ *satisfies*

$$(2-\kappa)\|\delta_0\| \leqslant \|\hat{x} - x_0\| \leqslant \kappa\|\delta_0\|.$$

The inclusion region $\overline{S(x_0, r)}$ of the Kantorovich theorem easily follows for the choice of scaled l_∞-norm and $A = F'(x_0)$.

4. Monotone convergence

The Newton method exhibits monotone convergence under partial ordering. We use the natural partial ordering for vectors and matrices; that is, $A \leqslant B$ $(A, B \in \mathbb{R}^{n\times n})$ if and only if $a_{ij} \leqslant b_{ij}$ $(i, j = 1, \ldots, n)$. The function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *convex* on a convex set $D \subseteq \mathbb{R}^n$ if

$$F(\lambda x + (1-\lambda)y) \leqslant \lambda F(x) + (1-\lambda)F(y) \tag{9}$$

holds for all $x, y \in D$ and $\lambda \in [0, 1]$. Assume that $F : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable on the convex set D. Then F is convex on D if and only if

$$F(y) - F(x) \geqslant F'(x)(y - x) \tag{10}$$

holds for all $x, y \in D$. The following basic result of Baluev is a special case of Theorem 13.3.7 of [42].

Theorem 15 (Baluev [42]). *Assume that* $F : \mathbb{R}^n \to \mathbb{R}^n$ *is continuously differentiable and convex on all of* $\mathbb{R}^n$, *that* $F'(x)$ *is nonsingular and* $[F'(x)]^{-1} \geqslant 0$ *for all* $x \in \mathbb{R}^n$, *and that* $F(x) = 0$ *has a solution* x^*. *Then* x^* *is unique and the Newton iterates* x_k *converge to* x^* *for any* x_0. *Moreover* $x^* \leqslant x_{k+1} \leqslant x_k$ $(k = 1, 2, \ldots)$.

Note that Baluev's theorem guarantees global convergence for the given function class. Characterizations of such functions and related results can be found in [42], or [18]. We show later that Newton's method has optimal complexity under special circumstances. It is not the case however for this kind of monotone convergence. Frommer [17] proved that Brown's method is faster than Newton's method under partial ordering. A similar result was proved for the ABS methods, as well [19].

5. The Newton method for underdetermined equations

For underdetermined equations of the form

$$F(x) = 0 \quad (F : \mathbb{R}^n \to \mathbb{R}^m,\ n \geqslant m) \tag{11}$$

an extension of the Newton iteration requires the solution of

$$F(x) + F'(x)(x_+ - x) = 0. \tag{12}$$

If $n > m$, then the new approximation x_+ is not uniquely determined. There are various ways to define x_+. The first result of this kind is due to Ben-Israel [6] who used the Moore–Penrose inverse to define $x_+ = x - F'(x)^+ F(x)$.

Theorem 16 (Ben-Israel [6]). *Let $F:\mathbb{R}^n \to \mathbb{R}^m$ be a function, $x_0 \in \mathbb{R}^n$, and $r>0$ be such that $F \in C^1(S(x_0,r))$. Let M, N be positive constants such that for all $x,y \in S(x_0,r)$ with $x-y \in \mathbb{R}\,(F'(y)^{\mathrm{T}})$:*

$$\|F(x)-F(y)-F'(y)(x-y)\| \leqslant M\|x-y\|,$$

$$\|(F'(x)^+ - F'(y)^+)F(y)\| \leqslant N\|x-y\|$$

and

$$M\|F'(x)^+\| + N = \gamma < 1 \quad (x \in S(x_0,r)),$$

$$\|F'(x_0)^+\|\|F(x_0)\| < (1-\gamma)r.$$

Then the sequence

$$x_{k+1} = x_k - F'(x_k)^+F(x_k) \quad (k=0,1,2,\ldots) \tag{13}$$

converges to a solution of $F'(x)^{\mathrm{T}}F(x)=0$ which lies in $S(x_0,r)$.

Condition $F'(x)^{\mathrm{T}}F(x)=0$ is equivalent to $F'(x)^+F(x)=0$. Algorithm (13) is called the normal flow algorithm. The name comes from the $n=m+1$ case, in which the iteration steps $-F'(x_k)^+F(x_k)$ are asymptotically normal to the Davidenko flow. For any n, the step $-F'(x_k)^+F(x_k)$ is normal to the manifold $\{y \in \mathbb{R}^n \mid F(y)=F(x_k)\}$. For the special case $n=m+1$ we refer to [2] and [10].

Walker [72] investigates the normal flow algorithm and the augmented Jacobian algorithm which is defined as follows. For a specified $V \in \mathbb{R}^{k\times n}$ and a given approximate solution x_k, determine x_{k+1} by

$$x_{k+1} = x_k + s, \text{ where } F'(x_k)s = -F(x_k) \text{ and } Vs = 0. \tag{14}$$

Walker gives two local convergence theorems under the following hypotheses.

Normal flow hypothesis. *F is differentiable, F' is of full rank m in an open convex set D, and the following hold*:

(i) *there exists $K \geqslant 0$* and $\alpha \in (0,1]$ *such that $\|F'(x)-F'(y)\| \leqslant K\|x-y\|^\alpha$ for all $x,y \in D$;*

(ii) *there is a constant B for which $\|F'(x)^+\| \leqslant B$ for all $x \in D$.*

Augmented Jacobian hypothesis. *F is differentiable and*

$$\begin{bmatrix} F'(x) \\ V \end{bmatrix}$$

is nonsingular in an open convex set D, and the following hold:

(i) *there exists $K \geqslant 0$ and $\alpha \in (0,1]$ such that $\|F'(x)-F'(y)\| \leqslant K\|x-y\|^\alpha$ for all $x,y \in D$;*

(ii) *there is a constant B for which*

$$\left\| \begin{bmatrix} F'(x) \\ V \end{bmatrix}^{-1} \right\| \leqslant B \quad \textit{for all } x \in D.$$

Furthermore for $\eta > 0$ let

$$D_\eta = \{x \in D \mid \|x - y\| < \eta \Rightarrow y \in D\}.$$

Theorem 17 (Walker [72]). *Let F satisfy the normal flow hypothesis and suppose D_η is given for some $\eta > 0$. Then there is an $\varepsilon > 0$ depending only on K, α, B, and η such that if $x_0 \in D_\eta$ and $\|F(x_0)\| < \varepsilon$, then the iterates $\{x_k\}_{k=0}^\infty$ determined by the normal flow algorithm* (13) *are well defined and converge to a point $x^* \in D$ such that $F(x^*) = 0$. Furthermore, there is a constant β for which*

$$\|x_{k+1} - x^*\| \leqslant \beta \|x_k - x^*\|^{1+\alpha}, \quad k = 0, 1, 2, \ldots .$$

Theorem 18 (Walker [72]). *Let F satisfy the augmented Jacobian hypothesis and suppose D_η is given for some $\eta > 0$. Then there is an $\varepsilon > 0$ depending only on K, α, B, and η such that if $x_0 \in D_\eta$ and $\|F(x_0)\| < \varepsilon$, then the iterates $\{x_k\}_{k=0}^\infty$ determined by the augmented Jacobian algorithm* (14) *are well defined and converge to a point $x^* \in D$ such that $F(x^*) = 0$. Furthermore, there is a constant β for which*

$$\|x_{k+1} - x^*\| \leqslant \beta \|x_k - x^*\|^{1+\alpha}, \quad k = 0, 1, 2, \ldots .$$

In fact, the latter theorem is a consequence of the previous one. The augmented Jacobian algorithm applied to F is equivalent to the normal flow algorithm applied to

$$\bar{F}(x) = \begin{bmatrix} F(x) \\ V(x - x_0) \end{bmatrix}.$$

The Moore–Penrose inverse is not the only possibility to define a Newton step in the underdetermined case. Nashed and Chen [39] suggested the use of outer inverses in a more general setting. Let X and Y be Banach spaces and let $L(X, Y)$ denote the set of all bounded linear operators on X into Y. Let $A \in L(X, Y)$. A linear operator $B : Y \to X$ is said to be an outer inverse, if $BAB = B$. The outer inverse of A will be denoted by $A^\#$. So the Newton method is given in the form

$$x_{k+1} = x_k - F'(x_k)^\# F(x_k), \quad k = 0, 1, 2, \ldots . \tag{15}$$

The following result is true.

Theorem 19 (Nashed–Chen [39]). *Let $F : D \subset X \to Y$ be Fréchet differentiable. Assume that there exist an open convex subset D_0 of D, $x_0 \in D_0$, a bounded outer inverse $F'(x_0)^\#$ of $F'(x_0)$ and constants $\eta, K > 0$, such that for all $x, y \in D_0$ the following conditions hold*:

$$\|F'(x_0)^\# F(x_0)\| \leqslant \eta,$$

$$\|F'(x_0)^\# (F'(x) - F'(y))\| \leqslant K \|x - y\|,$$

$$h := K\eta \leqslant \tfrac{1}{2}, \quad S(x_0, t^*) \subset D_0,$$

where $t^* = (1 - \sqrt{1 - 2h})/K$. Then (i) *the sequence $\{x_k\}$ defined by* (15) *with*

$$F'(x_k)^\# = [I + F'(x_0)^\# (F'(x_k) - F'(x_0))]^{-1} F'(x_0)^\# \tag{16}$$

lies in $S(x_0, t^*)$ *and converges to a solution* $x^* \in \overline{S(x_0, t^*)}$ *of* $F'(x_0)^\# F(x) = 0$; (ii) *the equation* $F'(x_0)^\# F(x) = 0$ *has a unique solution in*

$$\tilde{S} \cap \{R(F'(x_0)^\#) + x_0\},$$

where

$$\tilde{S} = \begin{cases} \overline{S(x_0, t^*)} \cap D_0 & \text{if } h = \frac{1}{2} \\ S(x_0, t^{**}) \cap D_0 & \text{if } h < \frac{1}{2} \end{cases}$$

$$R(F'(x_0)^\#) + x_0 = \{x + x_0 \mid x \in R(F'(x_0)^\#)\}$$

and

$$t^{**} = (1 + \sqrt{1 - 2h})/K;$$

(iii) *the speed of convergence is quadratic*:

$$\|x^* - x_{k+1}\| \leqslant \frac{1}{1 - Kt^*} \frac{K}{2} \|x^* - x_k\|^2 \quad (k = 0, 1, 2, \ldots).$$

For related results, we also refer to [9] where a generalization of Rall's Theorem 6 can be found. Finally we mention that Tapia [63] proved the convergence of the Newton method when the left inverse of $F'(x)$ is used. By extending the Gragg–Tapia results [20] Paardekooper [43] gave a Kantorovich-type inclusion region for the zero of $F(x) = 0$ $(F : X \to Y)$ when X and Y are Hilbert spaces, and the right inverse of $F'(x)$ is used.

6. Newton's method at singular points

Let X be a Banach space and $F : X \to X$. Assume that $F(x^*) = 0$ and the Jacobian $F'(x^*)$ is singular. The solution x^* is then called multiple or singular or non-isolated. Such situations may occur, for example, in the Bairstow method (see Blish–Curry [1, pp. 47–60]). The case of multiple zeros was first investigated by Rall [49]. Later Reddien [51] found a basic result which initiated an intensive research into singularities (see [22]). Here we recall only Reddien's result.

Assume that F is C^3 and $F'(x^*)$ has finite-dimensional null space N and closed range R so that $X = N \oplus R$. Let P_N denote a projection onto N parallel to R, and let $P_R = I - P_N$. The singular set of $F'(x)$ near x^* may range from a single point to a codimension one smooth manifold through x^*. Hence the nonsingularity of F' can be expected only in carefully selected regions about x^*. An added difficulty is that the Newton iterates must remain in the chosen region of invertability of F'. The following set satisfies both requirements:

$$W_{\rho,\theta} = \{x \in X \mid 0 < \|x - x^*\| \leqslant \rho, \|P_R(x - x^*)\| \leqslant \theta \|P_N(x - x^*)\|\}. \tag{17}$$

Theorem 20 (Reddien [52]). *Assume that*

(i) $\dim(N) = 1$;

(ii) $F''(x^*)(N, N) \cap R = \{0\}$;

(iii) *there is* $c > 0$ *so that for all* $\phi \in N$, $x \in X$, $\|F''(x^*)(\phi, x)\| \geqslant c\|\phi\| \, \|x\|$.

Then for ρ and θ sufficiently small, $F'(x)^{-1}$ exists for $x \in W_{\rho,\theta}$, the map $G(x)=x-F'(x)^{-1}F(x)$ takes $W_{\rho,\theta}$ onto itself, and there is $c_1 > 0$ such that $\|F'(x)^{-1}\| \leqslant c_1\|x-x^*\|^{-1}$ for all $x \in W_{\rho,\theta}$. Moreover if $x_0 \in W_{\rho,\theta}$ and $x_k = G(x_{k-1})$ for $k \geqslant 1$, the sequence x_k converges to x^* and the following hold:

$$\|P_R(x_k - x^*)\| \leqslant c_2\|x_{k-1} - x^*\|^2,$$

$$\lim_{k\to\infty} \|P_N(x_k - x^*)\|/\|P_N(x_{k-1} - x^*)\| = \tfrac{1}{2}.$$

Also, x^* is the only solution to equation $F(x)=0$ in the ball $\overline{S(x^*,\rho)}$.

Reddien's result indicates that the convergence region around x^* must have quite a special structure. Griewank [21] constructed an open star-like domain of initial points, from which the Newton method converges linearly to x^*. Griewank [22] provides a comprehensive survey of the singularity results.

7. The continuous Newton method

Gavurin [42] was the first to consider the continuous analogue of the Newton method

$$x'(t) = -[F'(x)]^{-1}F(x), \quad x(0) = x_0. \tag{18}$$

Let $x(t,x_0)$ denote the solution of (18) such that $x(0,x_0)=x_0$. We assume that $x(t,x_0)$ is defined on the maximum interval $[0,M)$. This solution satisfies the first integral

$$F(x(t,x_0)) = \exp(-t)F(x_0).$$

Hence, the image of the trajectory moves in the direction $F(x_0)$ towards the origin as time proceeds. Along this line the magnitude of $F(x)$ is reduced exponentially. If a solution exists for the interval $[0,\infty)$, then

$$\lim_{t\to\infty} F(x(t,x_0)) = 0.$$

Therefore we may expect that the solution approaches the set $V = \{x \mid F(x) = 0\}$. We also expect this behavior from the numerical solution of (18). If the explicit Euler method is applied on the grid

$$\{t_{k+1} \mid t_{k+1} = t_k + h_k,\ h_k > 0,\ k = 0,1,2,\ldots,\ t_0 = 0\},$$

we have the recursion

$$x_{k+1} = x_k - h_k[F'(x_k)]^{-1}F(x_k) \quad (k = 0,1,2,\ldots),$$

which becomes the "discrete" Newton method for $h_k = 1$ ($k \geqslant 0$). This is why the differential equation (18) is called the continuous or global Newton method. There are two questions:
(a) Under what conditions does $x(t,x_0)$ tend to a solution x^* of Eq. (2) as $t \to \infty$?
(b) What discretization method follows the solution path $x(t,x_0)$ to infinity?

Concerning question (a) we present the following results of Tanabe [61]. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable, $n \geqslant m$ and consider the continuous analogue of the Newton–Ben-Israel method

$$x'(t) = -[F'(x)]^{+}F(x), \quad x(0) = x_0, \tag{19}$$

where "+" stands for the Moore–Penrose inverse. Let $V_F = \{x \in \mathbb{R}^n \mid F(x) = 0\}$ and $S_F = \{x \in \mathbb{R}^n \mid \operatorname{rank}(F'(x)) = m\}$.

Theorem 21 (Tanabe [61]). *If* $\mathrm{rank}(F'(x^*)) = m$ *for a solution* $x^* \in V_F$ *then there exists a neighborhood* U^* *of* x^* *such that for each* $x_0 \in U^*$ *there exists a solution* $x(t,x_0)$, $0 \leqslant t < \infty$, *of* (19) *with* $x(0,x_0) = x_0$, *and as* t *tends to infinity it always converges to a point in* V_F *which may be different from* x^* *in the case where* $m < n$.

Theorem 22 (Tanabe [61]). *For a given* $x_0 \in S_F$, *there exists a solution* $x(t,x_0)$, $0 \leqslant t < M$, *of* (19) *with* $x(0,x_0) = x_0$. *As* t *tends to* M, *its trajectory either* (i) *converges to a solution* $x^* \in V_F \cap S_F$, *in which case we have* $M = \infty$ *and*

$$\|x(t,x_0) - x^*\| \leqslant k\|F(x_0)\| \exp(-t), \quad 0 \leqslant t < \infty$$

for some positive number k, *or* (ii) *approaches the set*

$$S = \{x \in \mathbb{R}^n \mid \mathrm{rank}(F'(x)) < m\}$$

of singular points of (19), *or* (iii) *diverges.*

The case $m > n$ is investigated in Tanabe [62]. We also mention that for $F : \mathbb{R}^n \to \mathbb{R}^n$ and a certain open bounded region $\Omega \subset \mathbb{R}^n$, Smale gave boundary conditions on $\partial\Omega$ under which the solution of (18) leads to a zero point x^* of F in Ω (see [2,23]).

Concerning question (b) we only note that it is not at all easy although the application of any sophisticated ODE solver seems straightforward. A list of papers dealing with the numerical implementation of the continuous Newton method is given in [62] (see also [2]). For derivation and theory of the continuous Newton method we refer to [2,11]. A quantitative analysis of the solution flow $x(t)$ of (18) in the presence of parameters is given in [25].

8. The Newton method for nonsmooth equations

Nonsmooth equations arise concerning various problems such as the nonlinear complementarity problem, variational inequalities and the Karush–Kuhn–Tucker system (see, e.g. [44,26]). The nonsmooth Newton methods are defined for functions $F : \mathbb{R}^n \to \mathbb{R}^m$ which are locally Lipschitz continuous. In such cases F is almost everywhere differentiable by Rademacher's theorem [15,54] and it is possible to use various kinds of generalized derivatives.

Suppose $F : \mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitzian function and let D_F denote the set of points at which F is differentiable. Let

$$\partial_B F(x) = \left\{ \lim_{x_i \to x,\ x_i \in D_F} F'(x_i) \right\}.$$

Let ∂F be the generalized Jacobian of F in the sense of Clarke [12]. Then $\partial F(x)$ is the convex hull of $\partial_B F(x)$,

$$\partial F(x) = \mathrm{conv}\, \partial_B F(x).$$

Let us denote

$$\partial_b F(x) = \partial_B F_1(x) \times \partial_B F_2(x) \times \cdots \times \partial_B F_m(x).$$

The classical directional derivative of F is defined by

$$F'(x;h) = \lim_{t\downarrow 0} \frac{F(x+th)-F(x)}{t}.$$

The function F is called semismooth at x if F is locally Lipschitzian at x and

$$\lim_{V\in\partial F(x+th'),\, h'\to h,\, t\downarrow 0} \{Vh'\}$$

exists for any $h \in \mathbb{R}^n$.

Suppose now that $F:\mathbb{R}^n \to \mathbb{R}^n$. The first nonsmooth Newton method is defined by

$$x_{k+1} = x_k - V_k^{-1}F(x_k) \quad (V_k \in \partial F(x_k),\ k = 0,1,2,\ldots). \tag{20}$$

An extension of the classical Newton–Kantorovich theorem is the following.

Theorem 23 (Qi–Sun [47]). *Suppose that F is locally Lipschitzian and semismooth on $\overline{S(x_0,r)}$. Also suppose that for any $V \in \partial F(x)$, $x, y \in \overline{S(x_0,r)}$, V is nonsingular,*

$$\|V^{-1}\| \leqslant \beta, \quad \|V(y-x) - F'(x;y-x)\| \leqslant K\|y-x\|,$$

$$\|F(y) - F(x) - F'(x;y-x)\| \leqslant \delta\|y-x\|,$$

where $q = \beta(\gamma+\delta) < 1$ and $\beta\|F(x_0)\| \leqslant r(1-q)$. Then the iterates (20) *remain in $\overline{S(x_0,r)}$ and converge to the unique solution x^* of $F(x)$ in $\overline{S(x_0,r)}$. Moreover, the error estimate*

$$\|x_k - x^*\| \leqslant \frac{q}{1-q}\|x_k - x_{k-1}\|$$

holds for $k = 1,2,\ldots$.

A modified nonsmooth Newton method is defined by

$$x_{k+1} = x_k - V_k^{-1}F(x_k) \quad (V_k \in \partial_B F(x_k),\ k = 0,1,2,\ldots). \tag{21}$$

Qi [48] proved the local superlinear convergence of this method. For the case $F:\mathbb{R}^n \to \mathbb{R}^m$ Chen et al. [9] suggested the following nonsmooth algorithm:

$$x_{k+1} = x_k - V_k^{\#}F(x_k) \quad (V_k \in \partial_B F(x_k),\ k = 0,1,2,\ldots), \tag{22}$$

where $V_k^{\#}$ denotes the outer inverse of V_k. Convergence results and numerical experiments can be found in [9].

Papers [44,26,9], contain further references to nonsmooth Newton papers.

9. The convergence and divergence of the Newton method

Under the standard assumptions the Newton method is locally convergent in a suitable sphere centered at the solution. We may ask however for the set of all points x_0 from which the Newton method is converging to a solution. The continuous Newton methods provide a possibility to characterize the set of convergence points. The case $\mathbb{R}^n$ is investigated in [2] (see also [62]). Braess [7] studies the case of complex polynomials. Another possibility is to use the results and techniques of iteration theory (see e.g. [65]). The best results in this direction are obtained for real and complex polynomials. The following observation indicates the difficulty of the convergence problem.

Theorem 24 (Rényi [52]). *Let $f:\mathbb{R}\to\mathbb{R}$ be defined on $(-\infty,+\infty)$. Let us suppose that $f''(x)$ is monotone increasing for all $x\in R$ and that $f(x)=0$ has exactly three real roots A_i $(i=1,2,3)$. The sequence $x_{k+1}=x_k-f(x_k)/f'(x_k)$ converges to one of the roots for every choice of x_0 except for x_0 belonging to an enumerable set E of singular points, which can be explicitly given. For any $\varepsilon>0$ there exists an interval $(t,t+\varepsilon)$ and in this interval three points a_i, $(t<a_i<t+\varepsilon,\ i=1,2,3)$ having the property that if $x_0=a_i$, $\{x_k\}$ converges to A_i $(i=1,2,3)$.*

The possibility that a small change in x_0 can cause a drastic change in convergence indicates the nasty nature of the convergence problem. The set of divergence points of the Newton method is best described for real polynomials.

Theorem 25 (Barna [5]). *If f is a real polynomial having all real roots and at least four distinct ones, then the set of initial values for which Newton's method does not yield a root of f is homeomorphic to a Cantor set. The set of exceptional initial values is of Lebesgue measure zero.*

Smale [58] gives a survey of results and related problems for complex polynomials. A geometric interpretation of the complex Newton method and its use for the convergence problem is given in [70] where a list of relevant publications is also given. The probability that the damped Newton method is converging to a zero is investigated in Smale [57] for complex polynomials.

10. Error analysis

Lancaster [32], Rokne [55] and Miel [35] investigated the following error propagation model of the Newton method:

$$\xi_{k+1}=\xi_k-[F'(\xi_k)+E_k]^{-1}(F(\xi_k)+e_k)+g_k \quad (k=0,1,2,\dots),$$

where E_k, e_k and g_k are perturbations and ξ_k is the computed Newton iterate instead of x_k. Under certain assumptions it is shown that the error sequence $\{\|x_k-\xi_k\|\}$ is bounded. If for some $k=p$ and some $l\geqslant 1$, $\xi_p=\xi_{p+l}$ and $x_k\to x^*$, then $\|\xi_k-x^*\|\leqslant\delta_0$ holds for $k\geqslant p$.

Wozniakowski [75] investigates the Newton method on the parametrized nonlinear system

$$F(x)=F(x;d)=0 \quad (F,x\in C^n,\ d\in C^m), \tag{23}$$

where vector d is the parameter. It is assumed that a simple zero x^* of (23) exists and F is sufficiently smooth in x and d. Let $\{x_k\}$ be a computed sequence of the successive approximations of x^* by an iteration ϕ. Let ζ be the relative computer precision in fl arithmetics. An iteration ϕ is called *numerically stable*, if

$$\overline{\lim_k}\,\|x_k-x^*\|\leqslant\zeta(k_1\|x^*\|+k_2\|F_x'(x^*;d)^{-1}F_d'(x^*;d)\|\,\|d\|)+\mathrm{O}(\zeta^2).$$

An iteration ϕ called *well behaved* if there exist $\{\delta x_k\}$ and $\{\delta d_k\}$ such that

$$\overline{\lim_k}\,\|F(x_k+\delta x_k;d+\delta d_k)\|=\mathrm{O}(\zeta^2)$$

and

$$\|\delta x_k\|\leqslant k_3\zeta\|x_k\|,\quad \|\delta d_k\|\leqslant k_4\zeta\|d\|$$

for large k. The k_i values can only depend on n and m ($i = 1, 2, 3, 4$). If ϕ is well behaved then it is also numerically stable. An algorithm of one Newton step in fl arithmetics is given by
(i) compute $F(x_k)$, $F'(x_k)$,
(ii) solve a linear system $F'(x_k)z_k = F(x_k)$,
(iii) set $x_{k+1} = x_k - z_k$.
Let us assume that F is computed by a well-behaved algorithm, that is

$$fl(F(x_k; d)) = (I + \Delta F_k)F(x_k + \Delta x_k; d + \Delta d_k) = F(x_k) + \delta F_k, \tag{24}$$

where $\|\Delta F_k\| \leqslant \zeta K_F$, $\|\Delta x_k\| \leqslant K_x\|x_k\|$, $\|\Delta d_k\| \leqslant K_d\|d\|$ and

$$\delta F_k = \Delta F_k F(x_k) + F_x'(x_k)\Delta x_k + F_d'(x_k)\Delta d_k + O(\zeta^2). \tag{25}$$

Further, let us assume that

$$fl(F'(x_k; d)) = F'(x_k) + \delta F_k', \quad \delta F_k' = O(\zeta). \tag{26}$$

This means that we do not need a well behaved algorithm for the evaluation of $F'(x_k)$. Finally, let us assume that a computed solution of the linear system $F'(x_k)z_k = F(x_k)$ satisfies

$$(F'(x_k) + \delta F_k' + E_k)z_k = F(x_k) + \delta F_k, \tag{27}$$

where $E_k = O(\zeta)$. Then a computed approximation x_{k+1} from $x_{k+1} = x_k - z_k$ satisfies

$$x_{k+1} = (I + \delta I_k)(x_k - z_k), \tag{28}$$

where δI_k is a diagonal matrix and $\|\delta I_k\| \leqslant C_1\zeta$, C_1 depends on the norm.

Theorem 26 (Wozniakowski [75]). *If* (24), (26) *and* (27) *hold, then the Newton iteration is well behaved. Specifically it produces a sequence* $\{x_k\}$ *such that*

$$\overline{\lim_k} \|F(x_{k+1} + \Delta x_k - \delta I_k x_k; d + \Delta d_k)\| = O(\zeta^2),$$

where Δx_k, δI_k *and* Δd_k *are defined by* (24) *and* (28).

A different error model is given by Spellucci [60].

11. Complexity results

The computational complexity of the Newton method was investigated by Kung and Traub [30], Traub and Wozniakowski [66,67]. Kung and Traub investigated real functions f and a class of rational two-evaluation iterations $x_{k+1} = \phi(f)(x_k)$ without memory given in the following form. Let U_0, U_1, U_2 and nonnegative integers h, t, independent of f such that $U_0 : \mathbb{R} \to \mathbb{R}$ is a rational function,

$$U_1(x, y) = \sum_0^l a_i(x)y^i, \tag{29}$$

where $a_i : \mathbb{R} \to \mathbb{R}$ is a rational function, and

$$U_2(x, y, z) = x + \frac{\sum_0^m b_{i,j}(x)y^i z^j}{\sum_0^q c_{i,j}(x)y^i z^j}, \tag{30}$$

where $b_{i,j}, c_{i,j} : \mathbb{R} \to \mathbb{R}$ are rational functions. Then

$$\phi(f)(x) = U_2(x, f^{(h)}(z_0), f^{(t)}(z_1)), \tag{31}$$

where

$$z_0 = U_0(x), \quad z_1 = U_1(x, f^{(h)}(z_0)). \tag{32}$$

Without loss of generality $U_0(x) = x$ and $h = 0$ can be assumed and thus the iteration function ϕ can be written in the form

$$\phi(f)(x) = x + \frac{\sum_0^m b_{i,j}(x) f^i(x)(f^{(t)}(z_1))^j}{\sum_0^q c_{i,j}(x) f^i(x)(f^{(t)}(z_1))^j}. \tag{33}$$

Kung and Traub defined the efficiency measure of an iteration ϕ by

$$e(\phi, f) = \frac{\log_2 p(\phi)}{v(\phi, f) + a(\phi)},$$

where $p(\phi)$ is the order of convergence of ϕ, $v(\phi, f)$ is the evaluation cost and $a(\phi)$ is the combinatory cost. The cost is the number of arithmetic operations. Let $E_2(f)$ denote the optimal efficiency achievable by a rational two-evaluation iteration without memory. Kung and Traub [30] showed that

$$E_2(f) = \max\left(\frac{1}{c(f) + c(f') + 2}, \frac{1}{2c(f) + 5}\right),$$

where $c(f)$ and $c(f')$ are the cost of evaluating f and f', respectively. Depending on the relative cost of evaluation f or f', the optimal efficiency $E_2(f)$ is achieved by either the Newton iteration

$$\gamma(f)(x) = x - \frac{f(x)}{f'(x)},$$

or the Steffensen iteration ψ,

$$\psi(f)(x) = x - \frac{f^2(x)}{f(x + f(x)) - f(x)}.$$

The result of Kung and Traub shed new lights on the intrinsic values of the Newton method. For Banach spaces the complexity of the Newton method is investigated in [66,67].

Acknowledgements

The author wishes to thank Prof. T. Yamamoto for pointing out references [73,74], and Prof. E. Spedicato for his comments. The author is also indebted to the unknown referees for their several useful remarks which helped him to improve the paper.

References

[1] E.L. Allgower, K. Georg (Eds.), Computational Solution of Nonlinear Systems of Equations, Lecture Notes in Applied Mathematics, Vol. 26, AMS, Providence, RI, 1990.
[2] E.L. Allgower, K. Georg, Numerical Continuation Methods, Springer, Berlin, 1990.

[3] I.K. Argyros, On Newton's method and nondiscrete mathematical induction, Bull. Austral. Math. Soc. 38 (1988) 131–140.
[4] I.K. Argyros, On Newton's method under mild differentiability conditions and applications, Appl. Math. Comput. 102 (1999) 177–183.
[5] B. Barna, Über die Divergenzpunkte des Newtonschen Verfahrens zur Bestimmung von Wurzeln algebraischer Gleichungen, I, Publ. Math. Debrecen 3 (1953) 109–118; II, 4 (1956) 384–397; III, 8 (1961) 193–207; IV, 14 (1967) 91–97.
[6] A. Ben-Israel, A Newton–Raphson method for the solution of systems of equations, J. Math. Anal. Appl. 15 (1966) 243–252.
[7] D. Braess, Über die Einzugsbereiche der Nullstellen von Polynomen beim Newton–Verfahren, Numer. Math. 29 (1977) 123–132.
[8] P.N. Brown, Y. Saad, Hybrid Krylov methods for nonlinear systems of equations, SIAM J. Sci. Statist. Comput. 11 (1990) 450–481.
[9] X. Chen, Z. Nashed, L. Qi, Convergence of Newton's method for singular smooth and nonsmooth equations using adaptive outer inverses, SIAM J. Optim. 7 (1997) 445–462.
[10] M.T. Chu, On a numerical treatment for the curve-tracing of the homotopy method, Numer. Math. 42 (1983) 323–329.
[11] M.T. Chu, On the continuous realization of iterative processes, SIAM Rev. 30 (1988) 375–387.
[12] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[13] J.R. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[14] P. Deuflhard, G. Heindl, Affine invariant convergence theorems for Newton's method and extensions to related methods, SIAM J. Numer. Anal. 16 (1979) 1–10.
[15] H. Federer, Geometric Measure Theory, Springer, Berlin, 1969.
[16] I. Fenyő, Über die Lösung der im Banachsen Raume definierten nichtlinearen Gleichungen, Acta Math. Hungar. 5 (1954) 85–93.
[17] A. Frommer, Comparison of Brown's and Newton's method in the monotone case, Numer. Math. 52 (1988) 511–521.
[18] A. Frommer, Verallgemeinerte Diagonaldominanz bei nichtlinearer Funktionen I–II, Z. Angew. Math. Mech. 72 (1992) 431–444, 591–605.
[19] A. Galántai, The global convergence of the ABS methods for a class of nonlinear problems, Optim. Methods Software 4 (1995) 283–295.
[20] W.B. Gragg, R.A. Tapia, Optimal error bounds for the Newton–Kantorovich theorem, SIAM J. Numer. Anal. 11 (1974) 10–13.
[21] A.O. Griewank, Starlike domains of convergence for Newton's method at singularities, Numer. Math. 35 (1980) 95–111.
[22] A.O. Griewank, On solving nonlinear equations with simple singularities or nearly singular solutions, SIAM Rev. 27 (1985) 537–563.
[23] M.W. Hirsch, S. Smale, On algorithms for solving $f(x)=0$, Commun. Pure Appl. Math. 32 (1979) 281–312.
[24] B. Jankó, Rezolvarea Ecuatiilor Operationale Nelineare in Spatii Banach (Romanian), Editura Academiei RSP, Bucuresti, 1969.
[25] V. Janovský, V. Seige, Qualitative analysis of Newton's flow, SIAM J. Numer. Anal. 33 (1996) 2068–2097.
[26] H. Jiang, L. Qi, X. Chen, D. Sun, Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations, in: G. Di Pillo, F. Gianessi (Eds.), Nonlinear Optimization and Applications, Plenum Press, New York, 1996.
[27] L.V. Kantorovich, G.P. Akilov, Functional Analysis in Normed Spaces, Pergamon, Oxford, 1964.
[28] H.B. Keller, Newton's method under mild differentiability conditions, J. Comput. Systems Sci. 4 (1970) 15–28.
[29] C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations, SIAM, Philadelphia, 1995.
[30] K.T. Kung, J.F. Traub, Optimal order and efficiency for iterations with two evaluations, SIAM J. Numer. Anal. 13 (1976) 84–99.
[31] Lai, Hang-Chin, Wu, Pou-Yah, Error bounds of Newton type process on Banach spaces, Numer. Math. 39 (1982) 175–193.

[32] P. Lancaster, Error analysis for the Newton–Raphson method, Numer. Math. 9 (1966) 55–68.
[33] Yu.V. Lisenko, The convergence conditions of the Newton–Kantorovich method for nonlinear equations with Hölderian linearization (Russian), Dokl. Akad. Nauk Belarusi 38 (1994) 20–24.
[34] G.J. Miel, The Kantorovich theorem with optimal error bounds, Amer. Math. Monthly 86 (1979) 212–215.
[35] G.J. Miel, Unified error analysis for Newton-type methods, Numer. Math. 33 (1979) 391–396.
[36] G.J. Miel, Majorizing sequences and error bounds for iterative methods, Math. Comput. 34 (1980) 185–202.
[37] G.J. Miel, An updated version of the Kantorovich theorem for Newton's method, Computing 27 (1981) 237–244.
[38] I.P. Mysovskikh, On the convergence of L.V. Kantorovich's method of solution of functional equations and its applications (Russian), Dokl. Akad. Nauk SSSR 70 (1950) 565–568.
[39] M.Z. Nashed, X. Chen, Convergence of Newton-like methods for singular operator equations using outer inverses, Numer. Math. 66 (1993) 235–257.
[40] A. Neumaier, Simple bounds for zeros of systems of equations, in: R. Ansorge, Th. Meis, W. Törnig (Eds.), Iterative Solution of Nonlinear Systems of Equations, Lecture Notes in Mathematics, Vol. 953, Springer, Berlin, 1982, pp. 88–105.
[41] J.M. Ortega, The Newton–Kantorovich theorem, Amer. Math. Monthly 75 (1968) 658–660.
[42] J.M. Ortega, W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.
[43] M.H.C. Paardekooper, An upper and a lower bound for the distance of a manifold to a nearby point, J. Math. Anal. Appl. 150 (1990) 237–245.
[44] J. Pang, L. Qi, Nonsmooth equations: motivation and algorithms, SIAM J. Optim. 3 (1993) 443–465.
[45] F.A. Potra, V. Pták, Sharp error bounds for Newton's process, Numer. Math. 34 (1980) 63–72.
[46] V. Pták, The rate of convergence of Newton's process, Numer. Math. 25 (1976) 279–285.
[47] L. Qi, J. Sun, A nonsmooth version of Newton's method, Math. Programming 58 (1993) 353–367.
[48] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, Math. Oper. Res. 18 (1993) 227–244.
[49] L.B. Rall, Convergence of the Newton process to multiple solutions, Numer. Math. 9 (1966) 23–37.
[50] L.B. Rall, A note on the convergence of Newton's method, SIAM J. Numer. Anal. 11 (1974) 34–36.
[51] G.W. Reddien, On Newton's method for singular problems, SIAM J. Numer. Anal. 15 (1978) 993–996.
[52] A. Rényi, On Newton's method of approximation (Hungarian), Mat. Lapok 1 (1950) 278–293.
[53] W.C. Rheinboldt, On a theorem of S. Smale about Newton's method for analytic mappings, Appl. Math. Lett. 1 (1988) 69–72.
[54] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer, Berlin, 1998.
[55] J. Rokne, Newton's method under mild differentiability conditions with error analysis, Numer. Math. 18 (1972) 401–412.
[56] H. Schwetlick, Numerische Lösung nichtlinearer Gleichungen, VEB DVW, Berlin, 1979.
[57] S. Smale, The fundamental theorem of algebra and complexity theory, Bull. Amer. Math. Soc. 4 (1981) 1–36.
[58] S. Smale, On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc. 13 (1985) 87–121.
[59] S. Smale, Newton's method estimates from data at one point, in: R.E. Ewing, K.I. Gross, C.F. Clyde (Eds.), The Merging Disciplines: New Directions in Pure, Applied, and Computational Mathematics, Springer, New York, 1986, pp. 185–196.
[60] P. Spellucci, An approach to backward analysis for linear and nonlinear iterative methods, Computing 25 (1980) 269–282.
[61] K. Tanabe, Continuous Newton–Raphson method for solving an underdetermined systems of nonlinear equations, Nonlinear Anal. Theory Methods Appl. 3 (1979) 495–503.
[62] K. Tanabe, Global analysis of continuous analogues of the Levenberg–Marquardt and Newton–Raphson methods for solving nonlinear equations, Ann. Inst. Statist Math. Part B 37 (1985) 189–203.
[63] R.A. Tapia, The weak Newton method and boundary value problems, SIAM J. Numer. Anal. 6 (1969) 539–550.
[64] R.A. Tapia, The Kantorovich theorem for Newton's method, Amer. Math. Monthly 78 (1971) 389–392.
[65] G. Targonski, Topics in Iteration Theory, Vandenhoeck & Rupprecht, Göttingen, 1981.
[66] J.F. Traub, H. Wozniakowski, Convergence and complexity of Newton iteration for operator equations, J. Assoc. Comput. Mach. 26 (1979) 250–258.

[67] J.F. Traub, H. Wozniakowski, Convergence and complexity of interpolatory-Newton iteration in a Banach space, Comput. Math. Appl. 6 (1980) 385–400.
[68] T. Yamamoto, Error bounds for Newton's iterates derived from the Kantorovich theorem, Numer. Math. 48 (1986) 91–98.
[69] T. Yamamoto, A method for finding sharp error bounds for Newton's method under the Kantorovich assumptions, Numer. Math. 49 (1986) 203–220.
[70] L. Yau, A. Ben-Israel, The Newton and Halley methods for complex roots, Amer. Math. Monthly 105 (1998) 806–818.
[71] B.A. Vertgeim, On the conditions of Newton's approximation method, Dokl. Akad. Nauk SSSR 110 (1956) 719–722 (in Russian).
[72] H.F. Walker, Newton-like methods for underdetermined systems, in: E.L. Allgower, K. Georg (Eds.), Computational Solution of Nonlinear Systems of Equations, Lecture Notes in Applied Mathematics, Vol. 26, AMS, Providence, RI, 1990, pp. 679–699.
[73] X. Wang, D. Han, On dominating sequence method in the point estimate and Smale's theorem, Sci. Sinica Ser. A (1989) 905–913.
[74] D. Wang, F. Zhao, The theory of Smale's point estimation and its applications, J. Comput. Appl. Math. 60 (1995) 253–269.
[75] H. Wozniakowski, Numerical stability for solving nonlinear equations, Numer. Math. 27 (1977) 373–390.